View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Waseda University Repository

修 士 論 文 概 要 書

2010 年　　2 月提出

専攻名 （専門分野）	情報理工学	氏 名	石原　寛之	指　導 教　員	後藤　滋樹　印
研究指導名	後藤　滋樹	学籍番号	CD 5108B008-8		
研　究 題　目	DNS レコードおよびプレフィックスの特徴を用いたスパム検知法				

概要: インターネットの普及に伴い、電子メールはビジネスでも日常生活においても欠かせない存在になっている。しかし電子メールを利用するユーザに対して大量な spam メール(迷惑メール) が送られる問題がある。その結果として、ユーザに対して様々な弊害を生むんでいる。Message Labs によると、2009 年の年間スパムの平均レートでは全体のメールに対して 87.7 % が spam メールという結果が報告されている。また spam メールの内で 83.4 %がボットネットを起因とするものと発表されている。中でも専用のリソースを利用した spammer も見逃す事が決してできない。特に正常なメールサーバとして機能しているように装い、送信者認証などにも対応している spammer が現れており問題になっている。そこで本研究は大規模な電子メールのログ分析を通じて、ビジネスとして大量に spam メールを送信する spam メール事業者の存在を明らかにするとともに、そのような事業者のインフラの発見方法とスパムフィルタリングへの応用を提案する。

1 spam メール

受信者の意図を無視して送られて来る無差別な大量一括送信メールの事を指す。メール利用者への負担、サーバやネットワーク資源の浪費への影響がある。

2 SPF (Sender Policy Framework)

2.1 SPF (Sender Policy Framework) の概要

RFC4408 [1] で定められている送信者認証技術の 1 つである。電子メールに使用されるプロトコルである SMTP では spammer(迷惑メール送信者) が差出人アドレスの偽装を行い利用者に詐欺メールを送るという問題があった。この偽装の対策として SPF を用いると、送信者のメールアドレスのドメイン(@ 以降のアドレス部分) に関する偽装を検出する事が可能である。認証には DNS レコードを用いる。

2.2 SPF を利用した認証手順

以下に SPF を利用した認証手順を図 1 と共に以下の手順で説明する。

1. メール送信側では、送信元ドメインの DNS サーバで SPF レコードにメールサーバの IP アドレスを記述して公開する
2. 受信側のメールサーバは、メールの From からドメイン部を取り出して、その DNS サーバに SPF レコードの問い合わせを行う
3. SPF 記述情報と送信側メールサーバの情報が適合した場合に、信頼できる送信者だと判別される。

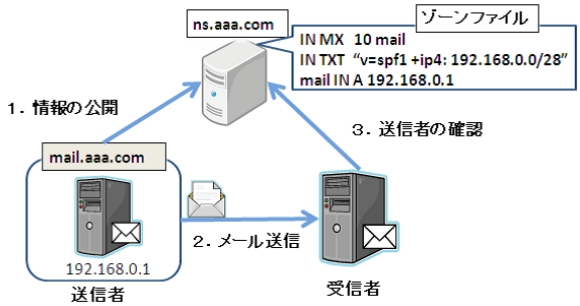

図 1: SPF を利用した認証手順

3 提案手法

3.1 提案手法の概要

専用のリソースを使用してspamメールを送るspammerは複数台のサーバを利用することが多い為、IPアドレスも同様にネットワーク単位で使用している可能性が高い。また送信者認証技術を利用して正常なシステムを装いspamメールを送ってくるspammerがある。そのようなspammerを検出する手法として本研究では以下の2通りの手法を提案する。

- 提案手法1: DNSレコードを用いたspammer検出
- 提案手法2: spammer prefixによるspammerの検出

3.2 提案手法1

送信者認証技術として導入されたSPF であるが、SPF を悪用する場合には、ドメインを取得した上でSPF を導入し正常なシステムとして振る舞うことで、送信者認証を欺く行為として使用されている。このような場合には、SPFレコード内に記述されたprefix (ネットワーク単位でのIPアドレス群) の多くがspamメールを送信しているメールサーバであると予測できる。

またSPFレコードだけではなく、MXレコードに関しても複数のメールサーバの設定が行われている可能性が高い。そこで提案手法1 ではsmtpログより取得したspamメールを送信してくるドメインに対してSPFレコードとMX レコードの参照を行い、そこに記述されているIPアドレスもしくはprefix に関してはspammerが送信を行っているものと推測する。このようにしてsmtpログから発見したspammerから未検知のspammer のインフラを発見する手法を提案する。

3.3 提案手法2

spammerが特定のネットワークから集中してメールを送信してくる事を意識して、すべてのネットワークを /29 のprefixとしてまとめ、spammerがprefix内で一定の割合で含まれている場合には、このprefix 内は spammer であると推定できる。

そこで提案手法2 としては、各prefix にspammerの送信元IPアドレスが含まれる割合を計算してspam prefixとして定義してspammerを検出する手法を提案する。

4 実証実験

4.1 実験に使用したデータ

実験には、ある企業網にて計測された 2009 年 3 月の 1 ヵ月間の smtp ログを使用する。この smtp ログには、メール送信者の IP アドレスとメールの分類が記載されている。メールの分類は以下の 3 種類である。

- spam: spam メールと判断された
- ham: 正常なメールと判断された
- grey list: grey list によって拒否された

本研究では spammer, legit の定義を以下に示す。

- spammer: spam メールを 9 割以上送っている。もしくは、grey list によって拒否されてメールの受信がない。
- legit: ham メールを 9 割以上送っている

smtp ログの spammer, legit, spam メールの数を表 1 に示す。

表 1: IP アドレスの総数と spammer, legit, spam メールの数

smtp ログ収集月	2009 年 3 月
IP アドレスの総数	1,148,559
spammer	1,022,038
legit	5,048
spam メール送信総数	91,589
grey list 総数	13,382,419

4.2 実験1

提案手法1を用いてspammer検出を行った。実験では、/16未満のprefixに関してはprefix値として大きすぎると判断した。spam prefix 作成後に、このspam prefixとsmtpログをフィルタリングして検出したsmtpログのスコアを参照して、spammer、legit、spamメール送信総回数、grey list総回数を調べた。また既存手法としてspamhaus.orgが提供するDNSBL を使用してフィルタリングを行った。結果を表2 に示す。提案手法1では、既存手法に比べてspammerの検出数が多い結果となった。legitの検出数は既存手法に比べるとおよそ15 倍の検出数となっておりfalse positiveが目立つ結果となっている。この原因は実験1ではprefixの値を/16未満のprefixに関しては除いたが、/16というサイズではprefix の単位としてはネットワークで考えるとかなり大規模なものであるため、このような結果になった可能性が高い。ただしspamメール送信総回数では提案手法では既存手法に比べておよそ20 倍のspamメールをブロックできる結果となってある。検出したspammerの1つ1つが大量のspamメールを送信しているspammerでありこのような悪意のあるspammerの検出ができることは有益な手法であると考えられる。

表 2: DNS レコードを用いた spammer 検出

	提案手法 1	既存手法
検出 IP アドレス数	5,760	1,599
spammer	1,381	437
legit	471	33
spam メール送信総回数	33,944	1,626
greylist 総回数	68,704	26,843

4.3 実験2

提案手法2を用いたspammer 検出を行った。prefixの中にspammerが含まれる割合に対してspam prefixの定義を変更した。各割合としては50%, 75%の2通りの基準でspam prefixを作成した。このspam prefixとsmtpログをフィルタリングして検出したsmtpログのスコアを参照して、spammer、legit、spamメール送信総回数、grey list総回数を調べた。結果を表3 に示す。prefixの中のspammer の割合が低くなればなるほど検出するspammer は多くなった。提案手法2における最大のメリットはlegit の数が表2 の既存手法と比べて小さく、spammerが50%の割合でspam prefixとしてもlegitの検出数は3という結果が得られた。この結果はspammerが特定のprefixに密集してspamメールを送信してきていることの裏付けるものである。またgrey listの総回数で比べると、/29のprefixでありながら、既存手法より44579回grey list検出を行っていることになる。つまり提案手法2で検出しているspammerの多くはgrey listで拒否された後に再送を行っていないspammerを多く検出できる事がわかる。

表 3: spammer prefix を用いた spammer 検出

	提案手法 2 50%	提案手法 2 75%
検出 IP アドレス数	1,590	499
spammer	1,239	454
legit	3	1
spam メール送信総数	569	131
grey list 総数	71,422	29,761

5 結論

本研究では、専用のリソースを確保して送信者認証を悪用し正常なシステムを装うことでspamメールを送信するspammerに対して、DNSレコードの情報を用いることで効率的にspammerのインフラを発見する手法と、spammerが一か所のネットワークの集中する特徴を利用してspammerが潜むネットワークを予測する手法を提案した。実験の結果、既存技術と比較してspammerの検出数が高い結果となった。spammer検出に関しては1つの手法が正しいというわけではなく、様々な方向よりspammerの特徴を見つけ出し評価を行っていく必要がある。

6 今後の課題

今後の課題として、本研究ではspammerのインフラよりISP毎にスコアをつけることで、個々のspammerのprefixから大規模なISP レベルまでの検出が行えるのではないかと考えられる。またspammerの定義さらに厳しくもしくは緩くすることでspammer検知の変化の調査を行い、より効果的にspammerを検出できる敷居値を見つける事が求められる。

参考文献

[1] M. Wong,W. Schlitt,”Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1”,RFC4408, April 2006.

[2] 本嶋悠也「IP アドレスの特徴を用いたspam メール判別方法」早稲田大学理工学部コンピュータ・ネットワーク工学科2008 年度卒業論文, 2008.